

VU Research Portal

Truth in Disagreement

Dumitrache, A.

2019

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Dumitrache, A. (2019). *Truth in Disagreement: Crowdsourcing Labeled Data for Natural Language Processing*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

CONTENTS

List of Figures xii

List of Tables xiv

1	INTRODUCTION	1
1.1	Motivation	1
1.2	Related Work	3
1.2.1	Crowd Data Quality	3
1.2.2	Crowdsourcing Aggregation Methods	5
1.2.3	Natural Language Processing with the Crowd	6
1.2.4	Capturing Ambiguity	6
1.3	Research Questions & Contributions	7
2	CROWDS VS. THE MEDICAL EXPERT	11
2.1	Introduction	11
2.2	Related Work	13
2.2.1	Medical Crowdsourcing	13
2.2.2	Crowdsourcing Ground Truth	14
2.2.3	Disagreement & Ambiguity in Crowdsourcing	14
2.3	Experimental Setup	15
2.3.1	Data Selection	15
2.3.2	CrowdTruth Metrics	17
2.3.3	Training the Model	19
2.3.4	Evaluation Data	20
2.3.5	CrowdTruth-Weighted Evaluation	21
2.4	Results	22
2.4.1	CrowdTruth vs. Medical Experts	22
2.4.2	CrowdTruth vs. Distant Supervision	25
2.5	Discussion	26
2.5.1	CrowdTruth vs. Medical Experts	26
2.5.2	CrowdTruth vs. Distant Supervision	28
2.5.3	Error Analysis & Limitations	29
2.6	Conclusion	29
2.7	Appendix: Evaluation Set Examples	31
3	DATA QUALITY FROM DISAGREEMENT	35
3.1	Introduction	35
3.2	CrowdTruth Methodology	37
3.2.1	CrowdTruth quality metrics	38
3.2.2	Spam Removal	40
3.3	Experimental Setup	41
3.3.1	Crowdsourcing Overview	42
3.3.2	Evaluation Methodology	46

3.3.3	Trusted Judgments Collection	47
3.4	Results	49
3.5	Discussion	51
3.5.1	CrowdTruth vs. Majority Vote	52
3.5.2	Finding the Right Number of Workers	53
3.6	Related Work	54
3.6.1	Crowdsourcing Ground Truth	54
3.6.2	Disagreement & Ambiguity in Crowdsourcing	55
3.6.3	Crowd Aggregation beyond Majority Vote	55
3.7	Conclusions	56
3.8	Appendix: Dataset Examples	58
4	LEARNING RELATION CLASSIFICATION FROM THE CROWD	61
4.1	Introduction	61
4.2	Related Work	63
4.3	Experimental Setup	64
4.3.1	Data and Crowdsourcing Setup	64
4.3.2	CrowdTruth Metrics	65
4.3.3	Label Propagation	66
4.3.4	Training the Relation Classification Model	67
4.4	Results and Discussion	67
4.4.1	Evaluating DS with CrowdTruth	68
4.4.2	Relation-Based Correction Experiment	70
4.4.3	Label Propagation Experiment	71
4.5	Conclusion	73
4.6	Appendix: Dataset Examples	74
5	FINDING AMBIGUITY FROM DISAGREEMENT	77
5.1	Introduction	78
5.2	Related Work	79
5.2.1	Crowdsourcing FrameNet	79
5.2.2	Disagreement & Ambiguity in Crowdsourcing	80
5.3	Crowdsourcing Setup	80
5.3.1	Dataset	80
5.3.2	Task Template	81
5.3.3	CrowdTruth Metrics for Capturing Disagreement	82
5.4	Crowd vs. Experts	83
5.5	Capturing Ambiguity	86
5.5.1	Ambiguity in the Frame-Sentence Expression	86
5.5.2	Ambiguity & Sentence Quality	88
5.5.3	Ambiguity & Frame Quality	89
5.6	A Frame Disambiguation Corpus with Ambiguity	91
5.6.1	Ambiguity in the Corpus	92
5.6.2	Systems Tested	93
5.6.3	Evaluation Metrics & Results	94
5.7	Conclusion	95
5.8	Appendix: Ambiguous Dataset Examples	96

6	CONCLUSION	99
6.1	Research Questions Revisited	99
6.2	Limitations & Future Directions	101
6.2.1	Disagreement beyond Relations & Frames	101
6.2.2	The Cost of Disagreement	102
6.2.3	Learning Ambiguity	103
	APPENDIX: CROWDTRUTH METRICS V.2.0	105
A	CrowdTruth Methodology	105
B	Building the Annotation Vectors	106
C	Disagreement Metrics	107
	BIBLIOGRAPHY	111
	SUMMARY	126
	SAMENVATTING	129
	REZUMAT	132
	ACKNOWLEDGEMENTS	136
	SIKS DISSERTATION SERIES	139